An Introduction to Set Theory

Fei Li[*]

March 1, 2016

In this article, we give a quick introduction to set theory. Section 1 contains a review of algebra of sets, functions, and power sets. Section 2 discusses Russell's paradox, following which we mention the axiom of specification in ZFC set theory. Sections 3 is on countability and uncountability of sets. We prove basic properties of countable sets, and then give a discussion of the Continuum Hypothesis.

1 Basic Notions

1.1 Sets, Algebra of Sets

A *set* is a collection of objects. Examples are:

1. $\{\text{Apple}, \text{Orange}, \text{Mango}\}$;
2. $\{1, 2, 3\}$;
3. $\{1, 2, 3, \{4\}\}$;
4. $\mathbb{N} = \{1, 2, 3, 4, 5, ...\}$, the set of natural numbers;
5. $\mathbb{R}$, the set of real numbers.

We write $u \in A$ if u is an element of A. We use $\varnothing$ to denote the *empty* set. The *union* of two sets A and B, denoted as $A \cup B$, is the set that in which we put all elements of A and B together. The *intersection* of two sets A and B, denoted as $A \cap B$, is the set that contains all common elements of A and B. See Figure 1.1.

$$A \cup B = \{u : u \in A \text{ or } u \in B\}; \tag{1}$$

$$A \cap B = \{u : u \in A \text{ and } u \in B\}. \tag{2}$$

Given two sets A and B, we say that A is a *subset* of B, denoted as $A \subseteq B$, if for every $u \in A$ we also have $u \in B$. The two sets are said to be *equal* if and only if $A \subseteq B$ and

[*] Email: fei.li.best@gmail.com. All rights reserved.

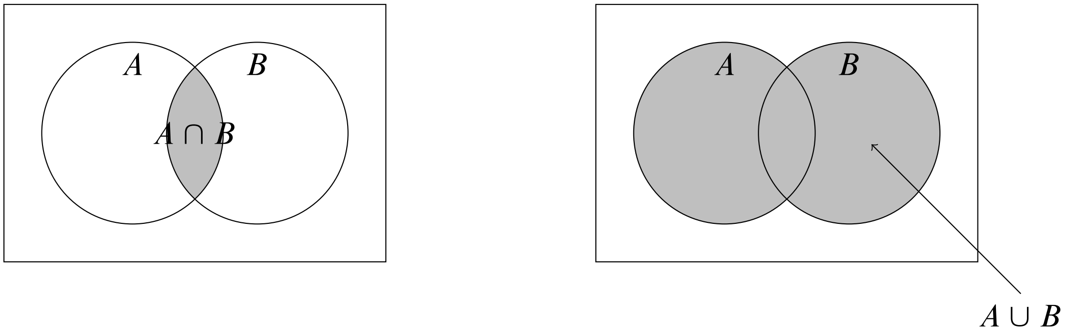


Figure 1.1: union and intersection of two sets

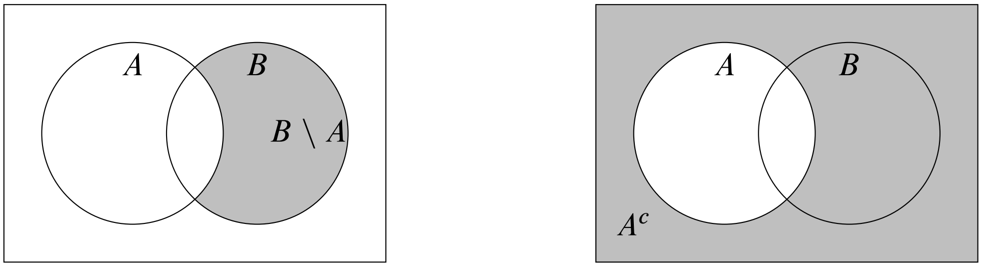


Figure 1.2: difference and complement of two sets

$B \subseteq A$, in which case we denote as $A = B$. Given A and B, $B \setminus A$, the *difference* of B to A is the set of all elements that is in B but not in A:

$$B \setminus A = \{x \in B : x \notin A\}. \tag{3}$$

Let $A \subseteq U$. The *complement* of A in U, denoted as A^c, is defined to be $U \setminus A$. See Figure 1.2. The following observations are obvious:

1. $A \cup \varnothing = A$; $A \cap \varnothing = \varnothing$;
2. $A \subseteq A$;
3. If $A \subseteq U$, then $A \cup U = U$, and $A \cap U = A$;
4. $A \cup A^c = U$, and $A \cap A^c = \varnothing$;
5. $(A^c)^c = A$.

Proposition 1.1 (De Morgan's laws)**.**

1. $(A \cup B)^c = A^c \cap B^c$*;*

2. $(A \cap B)^c = A^c \cup B^c$.

Proof. Let $u \in (A \cup B)^c$. Then u is excluded from both A and B. So on one hand u cannot be in A, that is, $u \in A^c$; on the other hand, u also cannot be in B, so $u \in B^c$. Thus, $u \in A^c \cap B^c$. This proves $(A \cup B)^c \subseteq A^c \cap B^c$.

To prove the reverse inclusion, let $u \in A^c \cap B^c$. If u is not in $(A \cup B)^c$, then it must be the case that $u \in A \cup B$. In this case, u is in either A or B. If $u \in A$, then we reach a contradiction: recall that our u is in A^c (as well as in B^c). Similarly, if $u \in B$, then this again contradicts our assumption about u. Thus $u \in (A \cup B)^c$, proving $(A \cup B)^c \supseteq A^c \cap B^c$. This completes the proof that $(A \cup B)^c = A^c \cap B^c$. The second relation is proved similarly. □

Given two sets A and B, their *Cartesian product* $A \times B$, is the set of all ordered pairs (a, b), where $a \in A$ and $b \in B$:

$$A \times B = \{\, (a, b) : a \in A \text{ and } b \in B\} \tag{4}$$

For example, if $A = \{1, 2, 3\}$, and $B = \{s, t\}$, then $A \times B = \{(1, s), (1, t), (2, s), (2, t), (3, s), (3, t)\}$. The set $\mathbb{R}^2 = \{(x_1, x_2) : x_1 \in \mathbb{R}, x_2 \in \mathbb{R}\}$ is the usual 2-dimensional Euclidean plane, and $\mathbb{R}^3 = \{\, (x_1, x_2, x_3) : x_i \in \mathbb{R} \text{ for each } i = 1, 2, 3\}$ is the 3-dimensional Euclidean space. Similarly, $\mathbb{R}^n$ is the set of all ordered pairs $(x_1, x_2, ..., x_n)$, where $x_i \in \mathbb{R}$ for each $i = 1, 2, ...n$.

1.2 Functions

Definition 1.2. Given two sets X and Y, a *function* $f : X \longrightarrow Y$ from X to Y associates each $x \in X$ with one element $f(x) \in Y$. X is called the *domain* of the function, and Y is called the *range* of the function.

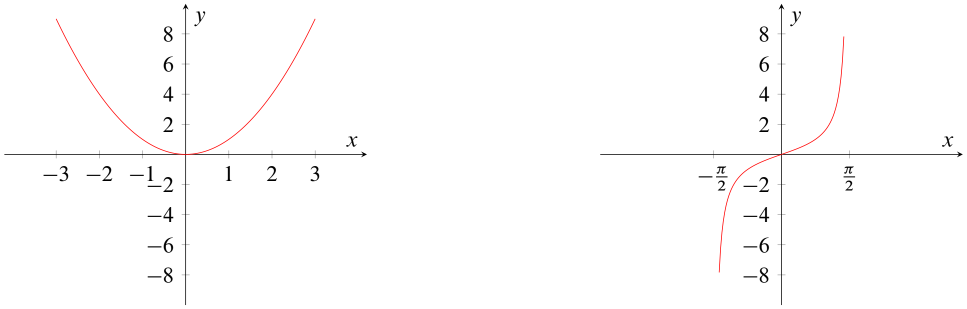
A function can send multiple elements in domain X to a single element y in range Y, but it cannot send an element $x \in X$ to multiple elements in Y. A function f is called *injective* (or *one-to one*) if for every $x_1 \neq x_2$ we have $f(x_1) \neq f(x_2)$. It is called *surjective* (or *onto*) if for every $y \in Y$ there is an $x \in X$ such that $f(x) \in Y$. It is called *bijective* (or *one-to-one and onto*) if it is both injective and surjective. A function may not be injective. $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is an example: $3 \neq -3$, but $3^2 = (-3)^2 = 9$. It is also not surjective: since squares of real numbers are nonnegative, no $x \in \mathbb{R}$ is sent to $(-\infty, 0)$. By contrast, $f : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ defined by $f(x) = \tan(x)$ is bijective. See Figure 1.3.

1.3 Power Sets

We use $|A|$ to denote the number of elements of a set A, i.e., the “cardinal" of A. For example, $|\{1, 2, 3\}| = 3$ since this set contains three elements.

Definition 1.3. For a set A, the *power set* of it is the set of all subsets of A. We denote it as $\mathscr{P}(A)$ or 2^A.

If $A = \{1, 2, 3\}$, then $\mathscr{P}(A) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. As you can see, the power set of A contains $8 = 2^3$ elements, i.e., $|\mathscr{P}(A)| = 2^{|A|}$. This is actually not a coincidence:



(a) $f(x) = x^2$ is neither injective nor surjective

(b) $f : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ defined by $f(x) = \tan(x)$ is bijective

Figure 1.3: injectiveness and surjectiveness of functions

Proposition 1.4. *For a finite set A, $|\mathscr{P}(A)| = 2^{|A|}$.*

Proof. To every subset of A we can associate a *characteristic function*. Specifically, if S is a subset of A, then we define

$$\mathbb{I}_S(u) = \begin{cases} 1 & \text{if } u \in S; \\ 0 & \text{if } u \notin S. \end{cases} \tag{5}$$

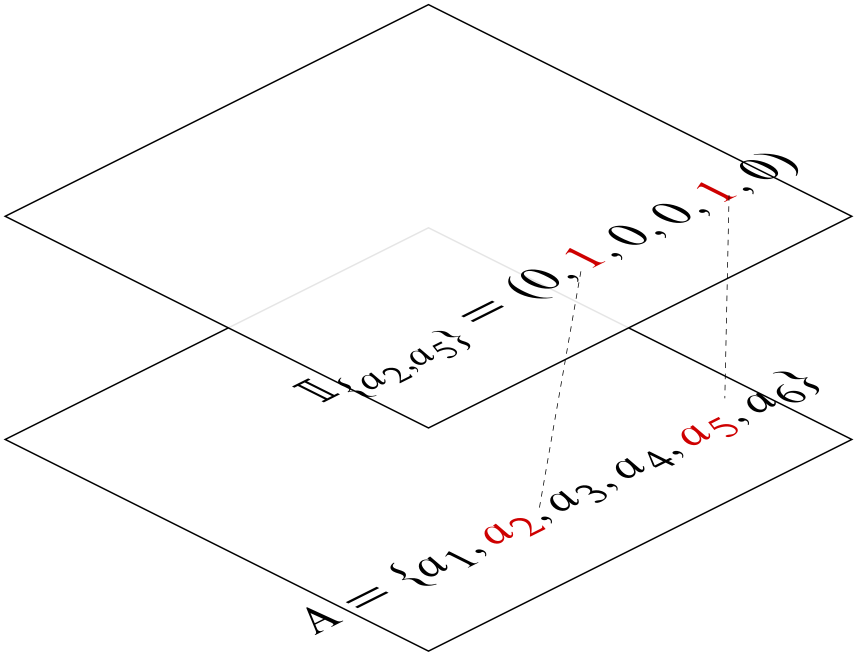
The characteristic function is an indicator of the subset: it “switches on" to 1 when an element $u \in A$ is in S, and it “switches off" to 0 when an element $u \in A$ is not in S. This function uniquely represents S, and different subsets have different characteristic functions associated with them. Thus, to count the total number of subsets of the set A, we need only count the number of all characteristic functions defined on A.

Suppose $|A| = n$ and write out the elements of A as $A = \{a_1, a_2, ..., a_n\}$. Then each element a_i of A is sent either to 0 or 1 by some characteristic functions. Thus each characteristic function can be writen out explicitly as an n-tuple. For example, $(1, 0, ..., 0)$ is the characteristic function on A that is equal to 1 on $\{a_1\}$ and 0 on the remaining part of A. Thus the function is $\mathbb{I}_{\{a_1\}}$ and this function represents the subset $\{a_1\}$. Similarly, $(1, 1, 0, ..., 0)$ is $\mathbb{I}_{\{a_1,a_2\}}$ and represents the subset $\{a_1, a_2\}$; $(1, 1, ..., 1)$ represents the whole set A, and $(0, 0, ..., 0)$ represents the empty set $\varnothing$. See Figure 1.4 for an illustration when A contains 6 elements.

So how many such n-tuples are there? The first slot can take two values, either 0 or 1. So every tuple is either $(0, * * *)$ or $(1, * * *)$. The second slot can also take either 0 or 1. So there are $2 \times 2 = 4$ cases regarding the first two slots: $(0, 0, * * *)$, $(0, 1, * * *)$, $(1, 0, * * *)$ and $(1, 1, * * *)$. Every tuple has one of the four cases. Upon each of the four cases, the third slot can also take 0 or 1, so there are $2 \times 2 \times 2 = 8$ cases of the first tree slots. Continuing this way, we see that the total number of such n-tuples must be

$$\underbrace{2 \times 2 \times ...... \times 2}_{n} = 2^n \tag{6}$$

This completes the proof. This proposition also explained the term “power set” as well as the notation 2^A. □



$\mathbb{1}_{\{a_2,a_5\}} = (0, 1, 0, 0, 1, 0)$
$A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$

Figure 1.4: the characteristic function of $\{a_2, a_5\}$ for $|A| = 6$

2 Russell's Paradox

We often create sets by specifying various conditions. For example, $\{x \in \mathbb{R} : x \geq 0\}$ is the set of nonnegative real numbers. If f and g are two real-valued functions defined on the interval $[0, 1]$, then $\{x \in [0, 1] : f(x) = g(x)\}$ is the subset of $[0, 1]$ on which the values of f and g coincide.

It seems that once we specify some arbitrary condition $P(x)$, the set $\{x : P(x)\}$ will always exist. In 1901, Bertrand Russell (1872–1970) showed that this can lead to paradox. Let the condition $P(x)$ be "$x \notin x$", i.e., "x is not a subset of itself". For example, if $x = \{$all triangles in the plane$\}$, then $x \notin x$, since the set itself is not a triangle. Denote this set by R:

$$R = \{x : x \notin x\}. \tag{7}$$

Then is $R \in R$? If $R \in R$, then R should satisfy the defining condition $P(x)$, which is "$x \notin x$", thus $R \notin R$; On the other hand, if $R \notin R$, i.e., if R is not a subset of itself, then R satisfy the condition "$x \notin x$", thus R is an element of the set R, i.e., $R \in R$. We are thus led to a paradox:

$$R \in R \Longrightarrow R \notin R; \tag{8}$$

$$R \notin R \Longrightarrow R \in R. \tag{9}$$

Conclusion. The set $R = \{x : x \notin x\}$ does not exist.

Russell's paradox has many real-life illustrations, one of them being the barber paradox. Suppose a barber only shaves $\{$*all men who do not shave themselves*$\}$. The question is, should the barber shave for himself? If he shaves himself, then he should fall into the set, i.e., he is a man who does not shave for himself. On the other hand, if he does not shave himself, then he is in the set {*all men who do not shave themselves*}. Since his job is to shave all the men in this set, he should also shave himself. The conclusion is that such a barber cannot exist.

Russell's paradox was proposed in the early days of set theory. It manifested the need for a system of *axioms* for the subject. Many axiomatic systems have been developed, but the standard and common one is the Zermelo–Fraenkel set theory, which, together with the Axiom of Choice, is abbreviated as ZFC. One of the axioms in Zermelo–Fraenkel set theory is the *axiom of specification*, which states:

Axiom of Specification. To every set A and every condition $P(x)$ there exists a subset B of A such that $B = \{x : P(x)\}$.

The axiom says that to search for the set that satisfies some condition $P(x)$, we have to first specify a place A, and our search result must lie in this place (the subset B in the axiom). We are not allowed to search for something out of nowhere, as in the case of Russell's paradox. Notice that both two examples given at the beginning of this section satisfy the axiom of specification: for the first one $A = \mathbb{R}$ and for the second one $A = [0,1]$. In contrast, the set $R = \{x : x \notin x\}$ does not specify a place for us to seek x according to $x \notin x$. The axiom of specification rules out such illusory sets, and the paradox is avoided.

3 Countability and Uncountability

Definition 3.1. For two sets S and T, we say that S *has cardinality less or equal to* T and denoted by $|S| \leq |T|$, if there is an injective function from S to T. We use $|S| < |T|$ to denote the situation where there exists an injective function from S to T, but no surjective function from S to T exists. We say that two sets S and T *have the same cardinality*, denoted by $|S| = |T|$, if there is a bijective function between S and T. If $|S| \leq |\mathbb{N}|$, then we say S is *countable*; otherwise ($|S| > |\mathbb{N}|$) it is said to be *uncountable*.

Theorem 3.2 (Schröder-Berstein Theorem)**.** *If* $|S| \leq |T|$ *and* $|T| \leq |S|$ *then* $|S| = |T|$.

Remark. The theorem says that if there exist an injective function from S to T, and an injective function from T to S, then there exists a bijection between S and T. While this seems intuitive, the proof of this theorem turn out to be a little bit complicated. The reason for the difficulty is that we are actually not comparing numbers, but we are seeking a bijective *function*, and it is not straightforward to construct such a function from two arbitrary injective functions.

Proof. Let f be an injective function from S to T, and let g be an injective function from T to S. We need to find a bijective function from S to T.

Define $S_0 := S \setminus g(T)$, and define $S_{n+1} = gf(S_n)$ recursively. Let $S_\infty = \bigcup_{n=1}^{\infty} S_n$. Define $h : S \to T$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \in S_\infty; \\ g^{-1}(x) & \text{otherwise.} \end{cases} \tag{10}$$

We claim that h is a bijective function from S to T.

We first verify h is injective. Given $x, y \in S, x \neq y$, we need to show that $h(x) \neq h(y)$. Now, if x and y are both in S_∞ or both in S_∞^c, then the injectiveness of h follows from the injectiveness of f and g. If $x \in S_\infty, y \in S_\infty^c$, and $h(x) = h(y)$, then $f(x) = g^{-1}(y)$ by the definition of h. But this means that $y = gf(x) \in gf(S_n) = S_{n+1}$ for some n, which contradicts our assumption about y. This shows that h is injective.

We next verify surjectiveness. Let $z \in T$. We need to find some element $x \in S$ such that $z = h(x)$. Let $x = g(z)$. If $x \notin S_\infty$, then $h(x) = g^{-1}(x) = z$. On the other hand, if $x \in S_\infty$, then $x \in S_n$ for some $n > 0$. By definition of S_n, this means that $x \in gf(S_{n-1})$, and consequently, $x = gf(x')$ for some $x' \in S_{n-1}$. Then $z = g^{-1}(x) = f(x') = h(x')$. This completes the proof that h is surjective. □

Every finite set is certainly countable; the natural number $\mathbb{N}$ itself is countable, yet it contains infinitely many elements. In this case we say that it is *countably infinite*. Now we prove some properties of countable sets.

Proposition 3.3. *If $|S| \leq |\mathbb{N}|$, then $|U| \leq |\mathbb{N}|$ for any $U \subseteq S$. (Any subset U of a countable set S is countable)*

Proof. This is obvious. Let f be an injective function from S to $\mathbb{N}$. Then f restricted to any subset U of S is an injective function from U to $\mathbb{N}$. Thus U is countable. □

Proposition 3.4. *Suppose $S = \bigcup_{i=1}^{\infty} S_i$, where $|S_i| \leq |\mathbb{N}|$ for each i. Then $|S| \leq |\mathbb{N}|$. (Countable union of countable sets are countable)*

Proof. List elements of each S_i as $\{a_{i1}, a_{i2}, a_{i3}, a_{i4}, ...\}$ and stack them together, we get the following array of S:

$$\begin{matrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

Label elements of S by $1, 2, 3, ...$ along the diagonal of the array as following, we obtain an injection from S to $\mathbb{N}$.

$$\begin{tikzcd}
a_{11} \arrow[d] & a_{12} \arrow[r] & a_{13} \arrow[dl] & a_{14} \arrow[r] & a_{15} \arrow[dl] \\
a_{21} \arrow[ur] & a_{22} \arrow[dl] & a_{23} \arrow[ur] & a_{24} \arrow[dl] & \\
a_{31} \arrow[d] & a_{32} \arrow[ur] & a_{33} \arrow[dl] & & \\
a_{41} \arrow[ur] & a_{42} \arrow[dl] & & & \\
a_{51} & & & & \\
\vdots & & & &
\end{tikzcd}$$

□

Corollary 3.5. $|\mathbb{Z}| = |\mathbb{N}|$. *(The set of integers $\mathbb{Z}$ is countable)*

Proof. Apply the above proposition to $Z = \{..., -3, -2, -1\} \cup \{0\} \cup \{1, 2, 3, ...\}$. □

Proposition 3.6. *Suppose $|S_1| \leq |\mathbb{N}|$ and $|S_2| \leq |\mathbb{N}|$. Then $|S_1 \times S_2| \leq |\mathbb{N}|$. (Cartesian product of two countable sets are countable)*

Proof. List elements of S_1 as $S_1 = \{a_1, a_2, a_3, ...\}$ and similarly list elements of S_2 as $S_2 = \{b_1, b_2, b_3, ...\}$. Then we can write out all elements of $S_1 \times S_2$ as

$$\begin{array}{ccccc}
(a_1, b_1) & (a_1, b_2) & (a_1, b_3) & (a_1, b_4) & \ldots \\
(a_2, b_1) & (a_2, b_2) & (a_2, b_3) & (a_2, b_4) & \ldots \\
(a_3, b_1) & (a_3, b_2) & (a_3, b_3) & (a_3, b_4) & \ldots \\
(a_4, b_1) & (a_4, b_2) & (a_4, b_3) & (a_4, b_4) & \ldots \\
(a_5, b_1) & (a_5, b_2) & (a_5, b_3) & (a_5, b_4) & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Now we can use the same counting method as in 3.4 to count the elements of $S_1 \times S_2$ along the diagonal. □

Corollary 3.7. $|\mathbb{Q}| = |\mathbb{N}|$. *(Rational numbers are countable)*

Proof. Recall that rational numbers $\mathbb{Q}$ is the set of numbers of the form $\frac{m}{n}$ with m and n relative prime to each other, like $\frac{2}{3}$ and $\frac{9}{1}$. We can write each element $\frac{m}{n}$ of $\mathbb{Q}$ as (m, n), thus making $\mathbb{Q}$ a subset of $\mathbb{Z} \times \mathbb{Z}$. Since $\mathbb{Z}$ is countable by 3.5, the above proposition implies that $\mathbb{Z} \times \mathbb{Z}$ is again countable. Then 3.3 implies that $|\mathbb{Q}| \leq |\mathbb{N}|$. On the other hand, the function $f : \mathbb{N} \to \mathbb{Q}$ defined by $f(n) = n$ for each $n \in \mathbb{N}$ is obviously injective, so $|\mathbb{N}| \leq |\mathbb{Q}|$. 3.2 now implies $|\mathbb{Q}| = |\mathbb{N}|$. □

Proposition 3.8. *If $|S| = |T|$, then $|\mathscr{P}(S)| = |\mathscr{P}(T)|$.*

Proof. Let f be a bijection between S and T. Then $F : \mathscr{P}(S) \to \mathscr{P}(T)$ defined by

$$F(A) = \{f(x) : x \in A\},\ A \subseteq S \tag{11}$$

is a bijection between $\mathscr{P}(S)$ and $\mathscr{P}(T)$. □

Remark. Surprisingly, the converse of the above theorem, i.e., $|\mathscr{P}(S)| = |\mathscr{P}(T)|$ impling $|S| = |T|$, is actually independent of ZFC.

Theorem 3.9. *For any set S, $|S| < |\mathscr{P}(S)|$.*

Proof. Suppose, by contradiction, that there exists a surjective function from S to $\mathscr{P}(S)$. Then, in particular, for the set

$$N = \{x \in S : x \notin f(x)\} \tag{12}$$

there should correspond to an element $x_0 \in S$ such that $f(x_0) = N$. Then is $x_0 \in N$? if $x_0 \in N$, then by the definition of N, $x_0 \notin f(x_0) = N$; On the other hand, if $x_0 \notin N = f(x_0)$, then again by definition $x_0 \in N$. We have showed

$$x_0 \in N \Rightarrow x_0 \notin N; \tag{13}$$

$$x_0 \notin N \Rightarrow x_0 \in N. \tag{14}$$

Thus there cannot exist a surjective function from S to the power set of S. □

In particular, $|\mathbb{N}| < |\mathscr{P}(\mathbb{N})|$, so $\mathscr{P}(\mathbb{N})$, the set of all subsets of $\mathbb{N}$, is uncountable. We now prove that there is actually a bijection between real numbers and $\mathscr{P}(\mathbb{N})$.

Theorem 3.10. $|\mathbb{R}| = |\mathscr{P}(\mathbb{N})|$.

Proof. We first find an injection from $\mathbb{R}$ to $\mathscr{P}(\mathbb{Q})$. For any real number r, the function

$$f(r) = \{q \in \mathbb{Q} : q < r\} \tag{15}$$

is injective from $\mathbb{R}$ to $\mathscr{P}(\mathbb{Q})$: if $r_1 \neq r_2$, then it is obvious that $f(r_1) \neq f(r_2)$, so $|\mathbb{R}| \leq |\mathscr{P}(\mathbb{Q})|$. Since $|\mathbb{Q}| = |\mathbb{N}|$ by 3.7, we have by 3.8 $|\mathscr{P}(\mathbb{Q})| = |\mathscr{P}(\mathbb{N})|$, so that $|\mathbb{R}| \leq |\mathscr{P}(\mathbb{N})|$.

We next prove $|\mathscr{P}(\mathbb{N})| \leq |\mathbb{R}|$. As in the proof of 1.4, we identify each subset of $\mathbb{N} = \{1, 2, 3, ...\}$ with its characteristic function, which in turn can be written out as an infinite sequence consisting of 0s and 1s. Then each sequence can be mapped to a real number r in $[0, 1)$, where r has its decimal expansion as the sequence, and the mapping is obviously injective. 3.2 then implies that $|\mathbb{R}| = |\mathscr{P}(\mathbb{N})|$. □

Figure 3.1: the Continuum Hypothesis states that there is a gap between discreteness and continuum.

Now we know that $|\mathbb{N}| < |\mathscr{P}(\mathbb{N})|$ by 3.9, and the later is actually the cardinality of the real numbers. $|\mathbb{N}|$ is the "number of elements" of a discrete infinite sequence, while $|\mathbb{R}|$ is the "number of elements" of a continuous infinite line. Then is there some other cardinals between them? In other word, is there some set S such that

$$|\mathbb{N}| < |S| < |\mathbb{R}|\ ? \tag{16}$$

Continuum Hypothesis (CH). No such S exists. In other word, if $|S| > |\mathbb{N}|$, then $|S| \geq |\mathbb{R}|$.

The Continuum Hypothesis is proven to be independent of ZFC by Kurt Gödel (1906–1978) and Paul Cohen (1934–2007). Note that discreteness is inherent in our definition of injectiveness and surjectiveness of functions, our notion of cardinality, and countability and uncountability of sets. After all, cardinality is about "counting the number of elements of a set", and counting is a "discrete" methodology. The Continuum Hypothesis reflects our inability to "count" uncountable sets: if some set is uncountable, then it contains "too many elements" to be distinguishable from the real line. In other words, the discrete concept of counting becomes less meaningful for continuums.